Check for updates

Human Factors, Volume 64, Issue 4, June 2022, pp. 662-674
DOI: 10.1177/0018720820960877

Penguins, Birds, and Pilot Knowledge: Can an Overlooked Attribute of Human Cognition Explain Our Most Puzzling Aircraft Accidents?

Richard Clewley and **Jim Nixon**, Cranfield University, Bedford, UK

Objective: We extend the theory of conceptual categories to flight safety events, to understand variations in pilot event knowledge.

Background: Experienced, highly trained pilots sometimes fail to recognize events, resulting in procedures not being followed, damaging safety. Recognition is supported by typical, representative members of a concept. Variations in typicality ("gradients") could explain variations in pilot knowledge, and hence recognition. The role of simulations and everyday flight operations in the acquisition of useful, flexible concepts is poorly understood. We illustrate uses of the theory in understanding the industry-wide problem of nontypical events.

Method: One hundred and eighteen airline pilots responded to scenario descriptions, rating them for typicality and indicating the source of their knowledge about each scenario.

Results: Significant variations in typicality in flight safety event concepts were found, along with key gradients that may influence pilot behavior. Some concepts were linked to knowledge gained in simulator encounters, while others were linked to real flight experience.

Conclusion: Explicit training of safety event concepts may be an important adjunct to what pilots may variably glean from simulator or operational flying experiences, and may result in more flexible recognition and improved response.

Application: Regulators, manufacturers, and training providers can apply these principles to develop new approaches to pilot training that better prepare pilots for event diversity.

Keywords: concepts, typicality, familiarity, training, knowledge, aviation safety

Address correspondence to Richard Clewley, Safety and Accident Investigation Centre, Cranfield University, Bedford MK43 0AL, UK; e-mail: r.m.clewley@cranfield.ac.uk

HUMAN FACTORS
Vol. 00, No. 0, Month XXXX, pp. 1-13
DOI:10.1177/0018720820960877
Article reuse guidelines: sagepub.com/journals-permissions
Copyright © 2020, The Author(s).

INTRODUCTION

In July 2013, a Boeing 777–200ER, operated by Asiana Airlines, struck the sea wall in front of the threshold of runway 28 Left at San Francisco International Airport, USA. The aircraft was destroyed by the impact and subsequent fire, and three passengers were fatally injured (National Transportation Safety Board [NTSB], 2014b). The crew were carrying out a visual approach, in good weather, but did not recognize unsafe aircraft energy and flight path conditions. For a sustained period, the aircraft was higher and faster than a typical approach, with a low thrust setting. At the final approach fix, just over five miles from touchdown, the aircraft was 450 feet too high, and shortly afterwards an incorrect autopilot mode was briefly selected, resulting in a pitch up command. The crew disconnected the autopilot, manually pitched down toward the runway, and selected idle thrust, but by 1000 feet, the approach path indications, visible from the cockpit, still indicated the aircraft was significantly above the desirable descent profile. As the aircraft passed through 400 feet, it briefly crossed the correct flightpath, but with idle thrust and a high descent rate the aircraft rapidly sank and the energy decayed below critical levels. Then, 2.5 s before ground impact, the pilot monitoring called "go-around," to abandon the landing, but it was too late, and the aircraft hit the sea wall and the tail section broke off. The aircraft never achieved the required stabilized approach criteria, exceeding flight path, thrust, speed, and vertical speed parameters.

There were over 26,000 flying hours in the cockpit, yet the crew were unable to recognize that the aircraft breached the approach safety criteria and therefore did not action the appropriate response protocol. This failure to recognize and respond to an unstabilized flight path on final approach is an enduring safety problem

(e.g., Air Accidents Investigation Branch [AAIB], 2014; Dutch Safety Board [DSB], 2010; NTSB, 2014a), but recognition failures have also featured in instrument data malfunctions (e.g., unreliable airspeed and stall; Bureau d'Enquêtes et d'Analyses [BEA], 2012) and flight control events (e.g., erroneous angle of attack data leading to activation of flight control augmentation system; Aircraft Accident Investigation Bureau of Ethiopia [AIB], 2019; Komite Nasional Keselamatan Transportasi [KNKT], 2019). How can experienced pilots, with many hours of simulator training, fail to recognize these events? Events that should be so familiar to knowledgeable flight crew. It is the aim of this study to answer that question.

In this study, we explore an overlooked aspect of human cognition: the theory of conceptual categories. This suggests humans reduce sensory experiences into groups that share features (Harnad, 2005; Pothos & Wills, 2011). Members of conceptual groups are not equal, and some prove more useful than others. For example, typical category members provide humans with a cognitive advantage, being easier to verify and learn (Rosch et al., 1976; Sandberg et al., 2012). This is the "typicality effect" (Rosch et al., 1976) and it is possible that exposure to typical event structures, in both simulated encounters and everyday work, may leave pilots with inadequate event knowledge, and poorly positioned to recognize and respond to some events (Clewley & Nixon, 2019). In this article, we review a literature that offers to explain how pilots might end up with significant variations in event knowledge, perhaps far from the robust, comprehensive knowledge we imagine. We find out where pilot knowledge is built and incubated, and examine the effect this may have on recognition. We develop two research hypotheses to locate systematic variations in pilot knowledge. We go on to discuss how this approach can help address weaknesses in pilot training and stimulate debate about how pilots acquire the sophisticated concepts required to manage some events.

THE THEORY OF CONCEPTUAL CATEGORIES

Conceptual categories are groups of objects and events that are treated equivalently by the cognitive system (Harnad, 2005). This means we can behave differently, to different things; a domestic cat and a tiger share many features, but our ability to recognize and discriminate is central to our cognitive and behavioral flexibility. There is renewed interest in conceptual categories (Hackett, 2017) and laboratory-based research has revealed a range of important knowledge effects that have not been exploited. There have been calls for its application to be widened to more real-world settings (Burnett et al., 2005; Goldstone et al., 2018). We respond to this and organize our research into two themes derived from literature on concepts: typicality and exemplar exposure.

Typicality

Typicality is the degree to which an instance serves as a good, central member of a concept (Rosch, 1975). Everyday concepts, such as "bird," exhibit variations in typicality—the robin is reliably judged more typical than the penguin (Nosofsky, 1988). Variations in typicality are described as typicality gradients. Rated typicality provides strong empirical evidence that concepts exhibit gradients (Dry & Storms, 2010). Typicality gradients have been observed in a variety of contexts, including animals and social situations (Dry & Storms, 2010). Typicality is cognitively advantageous. One such advantage, the typicality effect, suggests that typical instances are more readily learned and more easily recognized—they are stronger concepts and describe where our knowledge is concentrated (Rosch et al., 1976; Sandberg et al., 2012; Storms et al., 2000). The robin is more efficiently categorized as a bird than the penguin. It is a more representative and useful instance (Murphy & Ross, 2005).

The clearest and best cases of concept membership are known as prototypes (Rosch, 1975) and these act as cognitive reference points from which objects are judged more, or less, typical (Lakoff, 1987). Prototypical situations are more accessible, easier to describe, and richer in content, and allow actors to plan and order behavior (Cantor et al., 1982; Markman & Ross, 2003); this is a theoretical explanation of suboptimal behavior in nontypical events. Behavioral evidence of the typicality effect is supported

by electrophysiological data, which indicate typical items receive preferential processing, detectable in event-related potentials (Lei et al., 2010; Wang et al., 2016).

If typicality provides a cognitive advantage, then nontypical concepts may create cognitive disadvantage. This could cause delayed or inappropriate human response, which could represent risk to flight safety. This is a potential explanation for the difficulty in so-called reframing, when confronted with a surprising stimulus (Landman et al., 2017, 2018). The notion of typicality is compatible with sensemaking, and Weick (1995, p. 111) specifically notes that without "prototypical past moments" searching for context and meaning can be prolonged. Prototypical events are also found in the influential recognition-primed decision model, fostering more efficient strategies in decision-making environments like firefighting (Klein, 1998; Klein et al., 1986). Typical concepts feed into frames, and if knowledge is concentrated around a typical case, this will reduce cognitive flexibility when exotic events arise.

Typicality also accommodates the observation that some cases exhibit conceptual ambiguity. There are borderline, boundary cases that are not central or readily recognized because of conflicting or incomplete information (Genero & Cantor, 1987; Hampton, 1998). The aggregation of available cues and markers is insufficient for precise or adequate recognition (Clewley & Nixon, 2019). This feature is prominent in medical (mis)diagnosis and natural categories such as "birds" (e.g., featherless and flightless violating typical markers) reflecting diversity and variability within concepts. This principle has been linked to aircraft accidents (Clewley & Nixon, 2019). In 2009, the pilots of an Air France Airbus A330, operating flight AF447, were unable to recognize a loss of reliable airspeed data and an aerodynamic stall (BEA, 2012). Event ambiguity is reported as a significant factor in the history of AF447-type events. Thirteen other flight crews, from five different airlines, had experienced similar airspeed data malfunctions in flight, and at least nine of those crews had either not diagnosed or misdiagnosed the malfunction. None of the crews appeared to use the correct response protocol (BEA, 2012). Several crews misdiagnosed it as something similar, yet different, to the actual malfunction (BEA, 2012). This suggests these events were borderline, not immediately intelligible as a loss of air speed data, and not sufficiently clear to promote the use of an unusual memory checklist. This could suggest a systematic weakness in recognition capability, not errors specific to the crew of AF447. Typicality may be an important component of overall concept strength and unfavorable typicality gradients may be expressed through delayed or inappropriate pilot behavior.

We propose that typicality gradients drive pilot behavior. Knowledge is contoured and varied. The question of typicality is essentially how central a case is to a concept. This is the basis of any typicality hypothesis—a robin is more central to the concept "bird" than a penguin; a great white shark is more central to the concept "shark" than a hooded carpet shark; a chimpanzee is more central to the concept "primate" than a mongoose lemur. For this research, we have selected two event concepts from current trends in aircraft accidents after reviewing manufacturers' surveys and reports from three national accident investigation bodies. Using these event concepts, we address the first research hypothesis:

> *Hypothesis 1*: The candidate flight safety events will vary in how central they are to an event concept.

The events' "unstabilized approach to land" and "fuel system" were selected to test Hypothesis 1. Both event concepts feature in the Evidence-Based Recurrent Training Matrix for large public transport aircraft types (International Air Transport Association, 2013). Unstabilized approach to land events are flight path management competency based, and fuel system events are technical system management based. For each of these event concepts, three event types will be examined. The

development of the experimental stimuli is reported fully in the "Method" section.

Training With Exemplars: Where Aare Concepts Incubated?

Allied to typicality, familiarity is the subjective estimate of how often something has been experienced (Nosofsky, 1988). Familiarity covaries with repetition, so familiarity is an important driver of recognition, retrieval, and learning (Barsalou, 1985; Nosofsky, 1988; Rosch, 1978). Familiarity is associated with the exemplar theory of concepts, which proposes concept judgments are based on particular instances that have already been encountered (Smith & Medin, 1981), rather than prototypical reference points.

For pilots, familiarity can be built in either real-world encounters, or synthetic encounters, in a simulator. Understanding how familiarity relates to these contexts can indicate pilot exposure to event content. It is a proxy not only for experience, but also for the type and emphasis of experience. These experiences are likely to be the *only* direct opportunities for pilots to acquire crucial event knowledge that will inform recognition. The characteristics of each context, such as the diversity of exemplar and opportunity to acquire cues and markers, could explain why pilot knowledge later proves inadequate to recognize a real-world encounter.

We know of no data that specify patterns of pilot knowledge acquisition, yet these patterns could be important. Some event familiarity may primarily be built and incubated in simulated encounters that are predictable, brief, crude, or combined with other complex scenarios (Clewley & Nixon, 2019). The structure of airline pilot training means that abnormal events in the simulator are often expected. Training encounters may be predictable to the extent of negating any recognition benefits—pilots are so familiar with often-repeated events that they are not required to recognize them, but training may give the illusion of recognition (Casner et al., 2013). Pilot performance can drop below minimum acceptable standards even when well-understood events are presented in different contexts (Casner et al., 2013). Startle and surprise can be generated by including unexpected elements into simulator training, to boost pilot skills (Landman et al., 2018). Even so, current pilot training could pose a poorly understood, yet significant, burden on pilots to acquire event content amid cognitive constraints. A contrived, simple exemplar of a malfunction, perhaps involving highly sophisticated technology, may not provide adequate pilot knowledge. This is an "exemplar effect" (Clewley & Nixon, 2019), which may have influenced pilot behavior in the AF447-type events (BEA, 2012).

Conversely, everyday work in flight operations can be susceptible to repetition, and hence typicality effects. This could foster narrow concepts, like prototypes, especially if everyday work exhibits limited variety and diversity. This has the potential to reduce cognitive flexibility. When nontypical situations arise, such as unusual combinations of factors, pilots may be poorly positioned to recognize high-risk event markers. This is a candidate theoretical explanation for flight crew not recognizing unstabilized approach events. These events can carry risk of runway excursions and fatalities (e.g., NTSB, 2014a, 2014b), yet pilots sometimes continue approaches even with gross exceedances of parameters. If pilots rely on knowledge built in real-world flight operations, this could explain why they do not have the useful concepts needed to recognize unusual events.

We propose to examine these patterns of concept familiarity. Where is pilot knowledge built and incubated? Simulator experiences will provide exemplars of events, while flight operations will be subject to typicality effects that limit knowledge, revealing paucity or absence of training. An improved understanding of concept familiarity may explain some recognition failures and provide training insights. This leads to our second research hypothesis:

> *Hypothesis 2*: Pilot reliance on simulator or flight operations for event familiarity will vary for the candidate events.

METHOD

Design

We used a repeated-measures experimental design to test the hypotheses. Participants were required to respond to flight safety events built into scenario descriptions, similar in nature to clinical vignettes used in nursing research (e.g., Hughes & Huby, 2002). Scenarios were constructed with "slots" that accommodate the independent variable. The scenario descriptions are shown in the appendix.

Scenario Development

Event features were taken from accident reports in the public domain or extracted from Flight Crew Operating Manuals. A subject-matter expert in the host airline checked the scenarios for accuracy. Materials were piloted and minor adjustments were made so that the scenarios were representative of the aircraft type and flight operation at the host airline. Scenarios were developed around the following two themes.

Unstabilized approach to land events. This refers to a class of event where the aircraft fails to meet certain safety criteria when approaching to land (Clewley & Stupple, 2015). The tolerances are based on flight path (e.g., descending too quickly) or aircraft configuration (e.g., wheels not down). We used three types of specific event relating to the flight path parameters. This decision was based on the host airline's experience of these types of safety event. We represented all three flight path parameter event types at the host airline.

A "high speed" event refers to a speed tolerance being exceeded. A "high vertical speed" event refers to rate of descent exceedance. A "thrust idle" event refers to low engine thrust setting. In the scenario descriptions, one parameter exceeded tolerances and the other two parameters were normal. These events can lead to aircraft failing to meet landing performance criteria and runway excursions (e.g., AAIB, 2014).

Fuel system events. The fuel system refers to the storage and delivery of fuel from the tanks to the engines. We used three types of specific event. This was based on the host airline's fuel quantity status procedures from the Operations Manual.

A "minor fuel imbalance" refers to a condition where the fuel quantity in tanks no longer match. An "arrival fuel downward trend" refers to a condition where in-flight calculations show the predicted landing fuel to be diminishing, indicating the aircraft is using fuel at a higher rate than planned. A "fuel leak" refers to fuel that is escaping from the tank to engine closed system. These events can cause reduction in aircraft range and in-flight fires (e.g., AAIB, 2015).

Dependent Variables

We used a scale of 1–9 to measure typicality, using the anchors *not at all* (1) and *very* (9). These have been established and validated in previous research on concepts (e.g., Barsalou, 1987; Rothbart et al., 1996). A subject-matter expert at the host airline was consulted to provide options for source of concept familiarity. Four possibilities were identified: flight simulator devices, everyday flight operations, technical/operational manuals and checklists, and classroom study. Flight simulator devices and everyday flight operations provide direct event experience. Participants were asked to indicate the primary source of event knowledge.

Participants and Procedures

The research procedure and protocol were approved by the University Ethics Committee. Pilots were recruited at the host airline through workplace notices. After expressing interest, each respondent was sent a link to the Qualtrics survey platform (Qualtrics, Provo, Utah, USA) and completed the survey online.

In total, 118 pilots took part in the study. Two pools of pilots were used from two different bases. They all worked at the same European short-haul airline. Pool one: $N = 70$, 26 Captains, 44 First Officers, mean age 33.7 years (SD = 8.5), median flying experience 2800 hr, range = 13,900 hr, minimum 100 hr. Pool two: $N = 48$, 21 Captains, 27 First Officers, mean age 32.7 years, (SD = 7.5), median flying experience 2650 hr, range 10,820 hr, minimum 180 hr. All pilots were

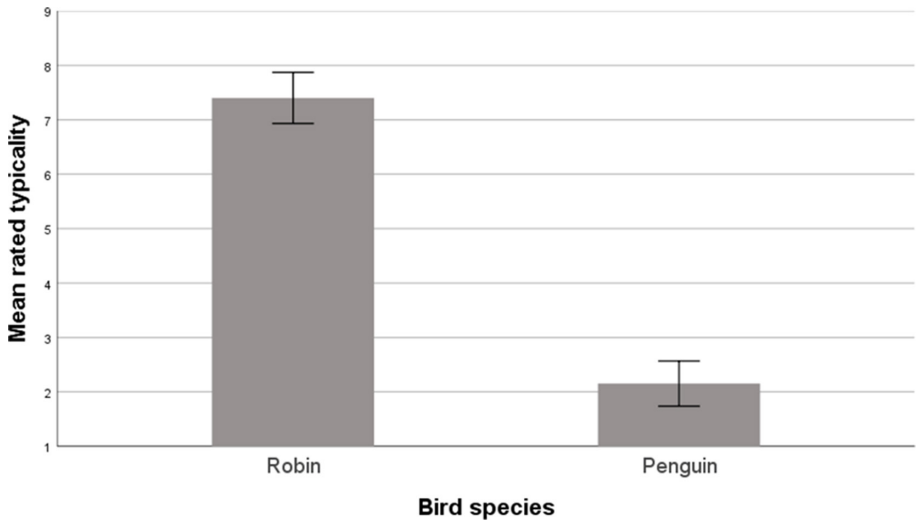


Figure 1. Mean rated typicality (with 95% confidence interval) for an everyday concept. This replicates a classic typicality gradient from concept research, indicating concentrations of knowledge and concept strength.

current in European short-haul flying operations on the same aircraft type.

Participants gave informed consent and completed the tasks online using the supplied link. The questions and response choices were presented in random order.

Data Analysis

Data analysis was carried out in IBM SPSS (version 25). We examined quartile-quartile plots for our typicality data, observing they approximated to normal distribution. We examined other literature on typicality ratings and found they exclusively use parametric statistical tests (e.g., Rosch et al., 1976; Rossiter & Best, 2013; Rothbart et al., 1996). We elected to use repeated-measure *t*-test and analysis of variance (ANOVA) to report the typicality gradients. There were no violations of sphericity. Pearson χ^2 was used to report main effects and post hoc differences in primary source of concept familiarity. Pairwise and post hoc comparisons were Bonferroni corrected. Effect size was reported as *r*, partial eta squared, or Cramer's *V*. The conventional α of <.05 was considered significant.

RESULTS

Typicality Ratings

To provide a baseline and replicate a classic typicality gradient, participants were asked to rate the everyday concept "bird." Figure 1 shows "robin" was rated as significantly more typical than "penguin," $t\,(47) = 12.65$, $p < .001$, $r = .84$.

Unstable approach to land events. Figure 2 shows mean rated typicality for the three approach types. There was a significant main effect for approach type, $F(1, 69) = 6.08$, $p < .005$, partial $\eta^2 = .081$. The high speed condition ($M = 5.76$, $SD = 2.01$, CI 95% [5.28–6.24]), was judged significantly more typical ($p < .05$) than the high vertical speed condition ($M = 4.91$, $SD = 2.26$, CI 95% [4.38–5.46]) and significantly more typical ($p < .01$) than the thrust idle condition ($M = 4.74$, $SD = 2.00$, CI 95% [4.26–5.22]). There was no significant difference of rated typicality between high vertical speed and thrust idle. Overall, the high-speed condition was rated significantly more typical than the high vertical speed and thrust idle conditions.

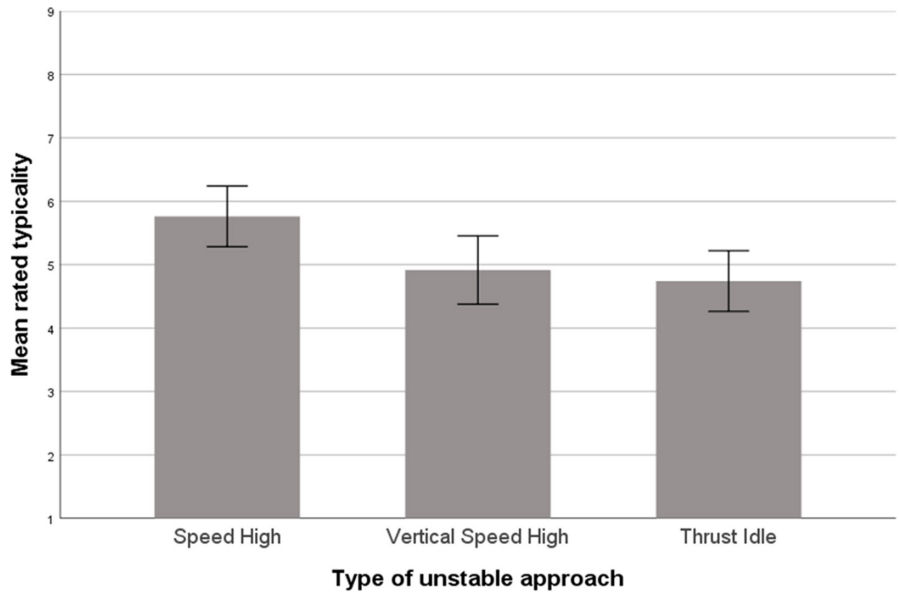


Figure 2. Mean rated typicality (with 95% confidence interval) for three types of unstable approach.

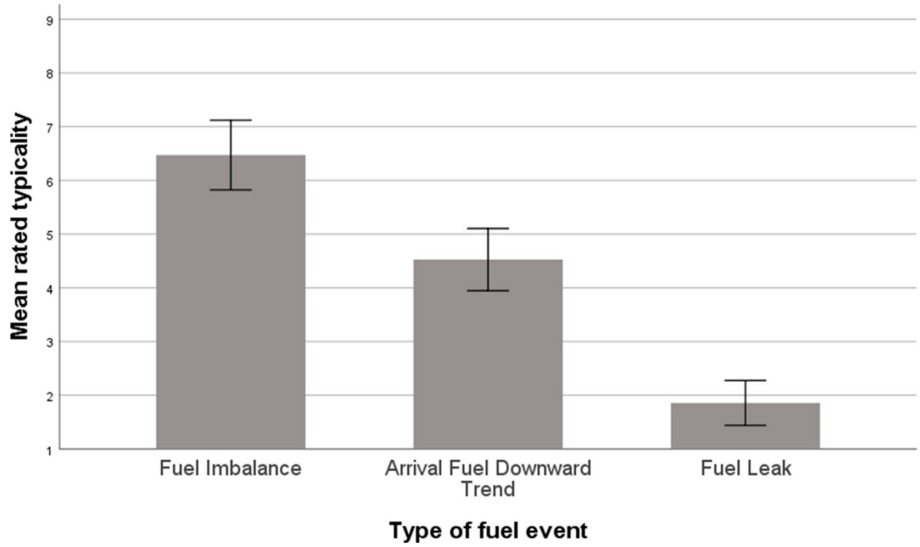


Figure 3. Mean rated typicality (with 95% confidence interval) for three fuel-related events.

Fuel system events. Figure 3 describes a typicality gradient across three fuel-related events.

There was a significant main effect for event type, $F(1, 47) = 66.94$, $p < .001$, partial $\eta^2 = .588$. The fuel imbalance ($M = 6.47$, $SD = 2.23$, CI 95% [5.82–7.13]) was judged significantly more typical ($p < .001$) than the arrival fuel downward trend ($M = 4.52$, $SD = 1.99$, CI 95% [3.95–5.10]) and the fuel leak ($M = 1.85$, $SD = 1.44$, CI 95% [1.44–2.74]). Pairwise comparison, with Bonferroni correction, indicates arrival fuel downward trend was judged

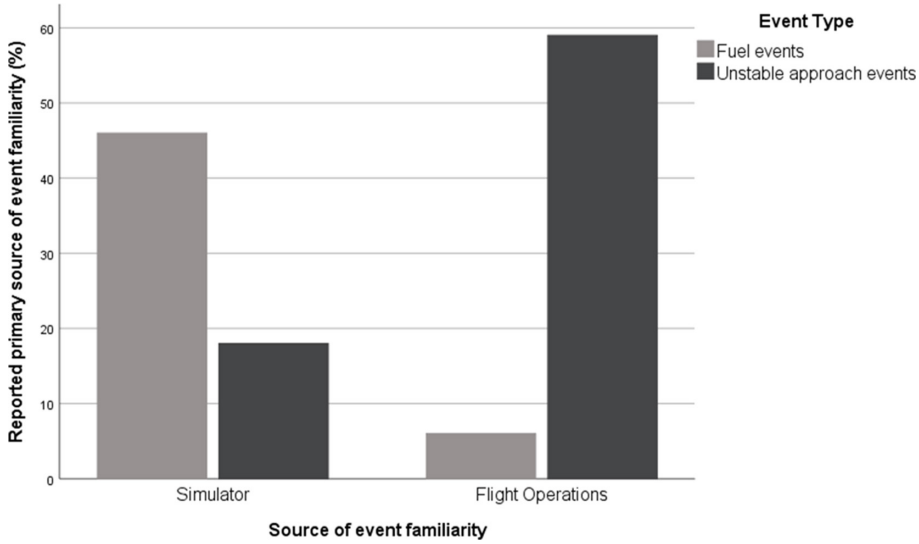


Figure 4. Comparing differences in reported primary source of event familiarity for simulator and flight operations.

significantly more typical than fuel leak ($p <$ .001).

Source of Concept Familiarity

There was a significant main effect for primary source of concept familiarity ($\chi^2(3) = 64.38$, $p < .001$, Cramer's V .567, $N = 70$). For fuel events, 46% of pilots reported the simulator as primary, 6% flight operations, 23% manuals, and 25% classroom. For unstable approach events, 18% reported the simulator as primary, 59% flight operations, 10% manuals, and 13% classroom. The post hoc comparison using adjusted residuals (Figure 4), with Bonferroni correction, indicates the flight simulator is more influential as a primary source of knowledge for fuel events, compared with unstable approach events ($\chi^2(1) = 18.01$, $p < .001$, $N = 70$). Flight operations are more influential as a primary source of knowledge for unstable approach events, compared with fuel events ($\chi^2(1) = 64.02$, $p < .001$, $N = 70$).

Overall, fuel events rely more heavily on simulated encounters when compared with unstable approach events. Unstable approach events exhibit a reversal of emphasis, placing more reliance on everyday flight operations.

DISCUSSION

We extended the theory of conceptual categories to flight safety events, translating two key elements, typicality and exemplar exposure, to a new domain. This approach provides new insights into the neglected topic of pilot knowledge and offers opportunities to address weaknesses in pilot training.

We presented evidence captured from airline pilots of significant typicality gradients and systematic variations in the source of event familiarity. Typicality gradients can locate concentrations of knowledge and concept strength that predict pilot performance. We found key gradients that may influence pilot behavior. Pilot knowledge is built and incubated in different psychological environments. These different environments may influence concept acquisition and the flexible deployment of knowledge in the real world. Some concepts may be over-reliant on simulated encounters, while others are linked to real flight experience, indicating paucity of training.

These findings provide empirical support to the lexicon and theoretical framework provided by Clewley and Nixon (2019), and stimulate debate about the industry-wide problem of nontypical flight safety events. Specifically, we developed

and addressed two research hypotheses that we discuss next.

H1: The Candidate Flight Safety Events Will Vary in How Central They Are to an Event Concept

We replicated a classic typicality gradient from literature and demonstrated that flight safety events exhibit the same gradients as everyday objects. We also demonstrated a significant typicality gradient for unstable approach to land events. The high-speed condition is rated as more typical than the high vertical speed and the thrust idle conditions. It is possible that such significant typicality advantages are important, but currently invisible, drivers of cognitive performance. Pilots may be more adept at recognizing and managing high-speed conditions because they are typical. For a given event, as pilots diverge from the typical case, cognitive performance declines (Rosch, 1975). Thus, cases proximal to the typical may be safer. They represent where knowledge is concentrated, just as our knowledge of birds is concentrated around typical species. This principle can now be applied to flight safety events to detect event types that warrant more specific, targeted training. Less typical forms of unstable approach, such as "thrust idle" conditions, are more likely to pose problems for pilot cognition, and several recent incidents show this is often not detected—a corollary of how the pilot monitoring task may also favor typicality (see DSB, 2010; NTSB, 2014b for examples of prolonged thrust idle conditions not detected). Typicality gradients locate candidate events subject to typicality effects.

The fuel events that we examined show a significant typicality gradient. This demonstrates a clear application of concept theory in identifying fragile concepts. Fuel leaks achieved the lowest mean typicality rating. This measurement is a proxy for concept strength. Our method could be applied to rapidly detect areas of weakness in concept knowledge to inform training and education. The fuel leak case may suffer poor recognition and response characteristics as pilots are likely to have limited knowledge of cues and response protocols, as well as limited direct experience (see later in the "Discussion" section for concepts built in simulated encounters). Locating weak domains of events could also lead manufacturers to providing flight crew with better in-flight documentation and improved guidance in checklists. This could deliver, and maintain, better connections between nontypical event concepts and the response protocol, as suggested by Clewley and Nixon (2019). This could also mitigate startle and surprise effects by providing better reframing knowledge for nontypical events (Landman et al., 2017).

H2: Pilot Reliance on Simulator or Flight Operations for Event Familiarity Will Vary for the Candidate Events

We found significant differences in the source of conceptual knowledge for our candidate events. This finding may be an important component in recognition failures by pilots. Two patterns emerged. The first pattern indicates pilot concepts for fuel system events are predominantly incubated in the simulator. The second pattern shows pilots rely on everyday flight operations to incubate and build conceptual knowledge of unstable approach events. This might indicate a lack of training for these event types and that the knowledge is simply built "on the job."

For fuel system events, concepts are likely to be incubated during simulated encounters. Encounters in everyday flight operations appear to have limited influence, perhaps because of system reliability. This places considerable burden on pilots during the simulated exemplars. Drawing on Casner et al. (2013), pilots must process event markers during simulator tests. There is scope to analyze whether pilots prioritize recognition skills or whether they are prevented from doing this by other cognitive tasks related to the simulation and skills test. If high-demand, nontypical events tend to be simulated exemplars, pilots should have adequate opportunity to process event content. It is significant that several other flight crews were unable to successfully recognize and manage an AF447-type airspeed sensor event (BEA, 2012). These reports from other flight crews suggest pilots may have weaker conceptual knowledge of these events. This could be explained by poor

opportunities to *acquire* markers in exemplars or that the real-world encounter is dissimilar to the simulator exemplar. It is possible that exemplars presented in training or testing cannot provide the sophisticated conceptual understanding needed in real world encounters.

Unstable approach familiarity favors everyday flight operations. This may indicate a lack of training for these events. Concepts crucial to recognition may have to be acquired on an ad-hoc basis in everyday work. This could explain why pilots have recognition problems. If everyday work exhibits typical patterns or limited variety, there may be little or no exposure to demanding sets of cues and markers. Perhaps these events are not amenable to high-quality simulations, so receive little useful attention in airline training regimes? This is a possible explanation for flight crews not recognizing unstable approaches (e.g., NTSB, 2014a, 2014b). The Asiana Airlines Boeing 777 accident, described in the opening paragraph, featured a combination of nontypical elements. Normal levels of automation were not used and normal approach aids were not available, providing key differences with typical task execution. This finding suggests we may not be equipping pilots with the flexible, useful concepts needed to recognize some events, perhaps simply due to paucity or absence of training.

Recommendations to Improve Safety

Explicitly train event concepts. Pilot training should train event concepts, embracing typical and nontypical event structures, building broader knowledge *around* the current simulator exemplars and everyday event prototypes. Do not just master an exemplar in the simulator and expect it to be a flexible knowledge tool. Likewise, everyday flight operations are not an ideal teacher of event concepts. Training the robin will not prepare us for the penguin. One way to achieve this is to provide innovative ways of transferring event content to pilots, such as in-cockpit digital applications. This could improve pilot knowledge of events like unstable approaches and nontypical technical events. We offer a theory-driven method to do that.

Locate key typicality gradients. Our results suggest pilots experience typicality gradients in flight operations that may damage event recognition. Our recommendation is to locate these gradients and understand the effect they have on pilot behavior and performance. For some event types and domains, the typicality gradients may be steep, giving pilots cognitive problems. Some aircraft technology can be poorly understood, even by training pilots (see NTSB, 2014b, for a discussion on pilot knowledge of the Boeing 777 auto flight behavior) so we think it is important to locate areas of weaker knowledge. This may indicate where to supplement pilot training and cockpit materials, for example, providing extra diagnostic guidance in checklists or adjusting memory-based protocols for nontypical events.

Strengths, Limitations, and Further Research

We have contributed a theory-driven analysis of pilot event knowledge adding typicality gradients, typicality effects, and exemplar effects to the human factors literature. We have proposed practical ways of linking this approach to innovations in training and pilot response protocols. This builds empirical evidence to support the theoretical apparatus proposed by Clewley and Nixon (2019) and provides an incremental contribution to a neglected area: pilot knowledge.

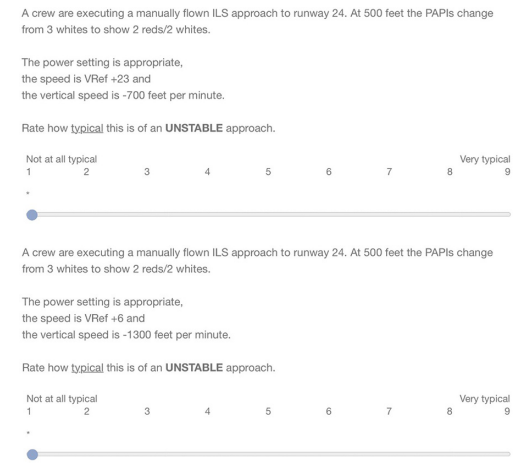
Naturally, our research has limitations. We have used two event types, although there is now scope to expand this to other domains of pilot event knowledge. Our event choices are not necessarily generalizable to the many other flight safety events that may have their own typicality profile. Nevertheless, we have shown that flight safety events are vulnerable to a typicality effect. Also, the type of flight operation may have influenced the results. Our sample were short-haul pilots. It is possible that long-haul pilots, spending more time in cruise phases and doing fewer landings, may have different experiences of events, such as fewer unstable approaches and greater reliance on simulated encounters. In our scenario descriptions, we were careful to use terminology derived from aircraft documentation, but participants were not immersed in a dynamic simulation or a real-world encounter. We suggest typicality gradients

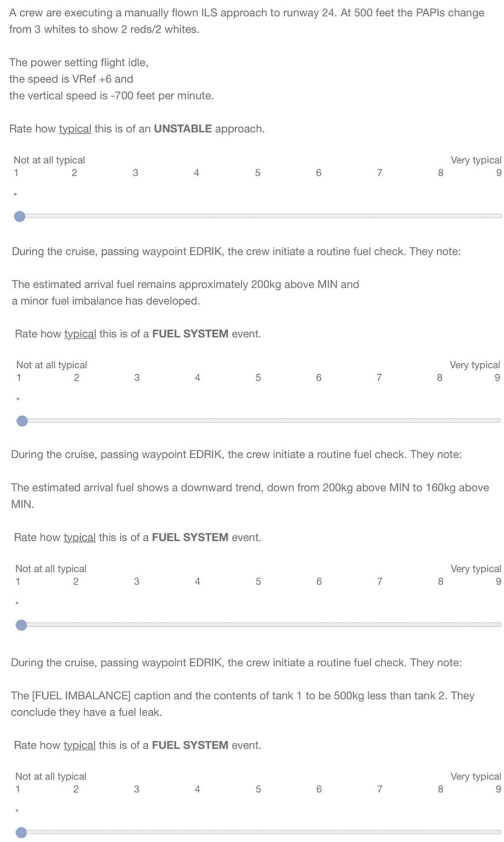
are explored with dynamic cues to further validate the approach. Real event encounters often evolve, so the availability of cues changes over time, and this may influence recognition. Future research should focus on measurement of cognitive performance predicted by typicality gradients to refine our understanding of the effects on recognition and response and elucidate the mechanisms involved—for example, clarifying the influence of typicality in extracting important perceptual event features and the complementary role of working memory.

There is reason to be optimistic that this theoretical approach could be aimed toward public emergency response events. For instance, in the United Kingdom in June 2017, a tower block fire caused significant loss of life and presented public emergency responders with a variety of complex problems (see Grenfell Tower Inquiry, 2017). These sorts of events are unusual and present an array of nontypical and unfamiliar stimuli that are demanding, even for experienced professionals. Concept theory is well positioned to expand explanations of these important phenomena and inform evidence-based improvements. This overlooked attribute of human cognition may explain some recent aircraft accidents, and, if better understood can be used to improve safety.

APPENDIX

Screenshots of All Six Scenario Descriptions Used in the Typicality Rating Task





ACKNOWLEDGMENTS

We thank the reviewers for their suggestions, particularly reviewer two, who suggested many helpful and insightful improvements. We also thank the Director of Flight Operations and Chief Pilot at the host airline for allowing us access to the pilot community.

KEY POINTS

- We demonstrate how concept theory can be applied to understand nontypical flight safety events.
- We provide empirical evidence of variations in typicality ("gradients") and systematic variations in the source of event familiarity.
- We explain why typicality and exemplar effects may degrade recognition, response and safety.
- We propose evolving pilot training and response protocols to explicitly train broader, more useful event concepts.

ORCID iDs

Richard Clewley https://orcid.org/0000-0002-3393-5991

Jim Nixon https://orcid.org/0000-0001-7072-6585

REFERENCES

Air Accidents Investigation Branch (AAIB). (2014). *AAIB Bulletin: 10/2014 VP-CKY.*

Air Accidents Investigation Branch (AAIB). (2015). Report on the accident to Airbus A319-131, G-EUOE London Heathrow Airport 24 May 2013.

Aircraft Accident Investigation Bureau of Ethiopia (AIB). (2019). *Aircraft accident investigation preliminary report: B737-8 (MAX) registered ET-AVJ 28 NM South East of Addis Ababa, Bole International Airport*.

Barsalou, L. W. (1985). Ideals, central tendency, and frequency of instantiation as determinants of graded structure in categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*, 629–654.

Barsalou, L. W. (1987). The instability of graded structure: Implications for the nature of concepts. In U. Neisser (Ed.), *Concepts and conceptual development: Ecological and intellectual factors in categorization* (pp. 101–140). Cambridge University Press.

Bureau d'Enquêtes et d'Analyses (BEA). (2012). *Final report on the accident 1st June 2009 to the Airbus A330-203 registered F-GZCP operated by Air France flight AF447 Rio de Janeiro – Paris*. BEA.

Burnett, R. C., Medin, D. L., Ross, N. O., & Blok, S. V. (2005). Ideal is typical. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *59*, 3–10. https://doi.org/10.1037/h0087453

Cantor, N., Mischel, W., & Schwartz, J. C. (1982). A prototype analysis of psychological situations. *Cognitive Psychology*, *14*, 45–77. https://doi.org/10.1016/0010-0285(82)90004-4

Casner, S. M., Geven, R. W., & Williams, K. T. (2013). The effectiveness of airline pilot training for abnormal events. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *55*, 477–485. https://doi.org/10.1177/0018720812466893

Clewley, R., & Nixon, J. (2019). Understanding pilot response to flight safety events using categorisation theory. *Theoretical Issues in Ergonomics Science*, *20*, 572–589. https://doi.org/10.1080/1463922X.2019.1574929

Clewley, R., & Stupple, E. J. N. (2015). The vulnerability of rules in complex work environments: Dynamism and uncertainty pose problems for cognition. *Ergonomics*, *58*, 935–941. https://doi.org/10.1080/00140139.2014.997804

Dry, M. J., & Storms, G. (2010). Features of graded category structure: Generalizing the family resemblance and polymorphous concept models. *Acta Psychologica*, *133*, 244–255. https://doi.org/10.1016/j.actpsy.2009.12.005

Dutch Safety Board (DSB. (2010). Crashed during approach, Boeing 737-800, near Amsterdam Schiphol Airport, 25 February 2009: Project number M2009LV0225_01.

Genero, N., & Cantor, N. (1987). Exemplar prototypes and clinical diagnosis: Toward a cognitive economy. *Journal of Social and Clinical Psychology*, *5*, 59–78. https://doi.org/10.1521/jscp.1987.5.1.59

Goldstone, R. L., Kersten, A., & Carvalho, P. F. (2018). Categorization and Concepts. In S. Thompson-Schil (Ed.), *Stevens' handbook of experimental psychology and cognitive neuroscience* (4th ed.,pp. 275–317). Wiley.

Grenfell Tower Inquiry. (2017). The Grenfell tower inquiry. Retrieved February 1, 2019, from https://www.grenfelltowerinquiry.org.uk/

Hackett, P. M. W. (2017). Editorial: Conceptual categories and the structure of reality: Theoretical and empirical approaches. *Frontiers in Psychology*, *8*, 1–2. https://doi.org/10.3389/fpsyg.2017.00601

Hampton, J. A. (1998). Similarity-based categorization and fuzziness of natural categories. *Cognition*, *65*, 137–165. https://doi.org/10.1016/S0010-0277(97)00042-5

Harnad, S. (2005). To cognize is to categorize: Cognition is categorization. In H. Cohen & C. Lefebvre (Eds.), *Handbook of categorization in cognitive science* (pp. 19–43). Elsevier.

Hughes, R., & Huby, M. (2002). The application of vignettes in social and nursing research. *Journal of Advanced Nursing*, *37*, 382–386. https://doi.org/10.1046/j.1365-2648.2002.02100.x

International Air Transport Association. (2013). *Evidence-based training implementation guide*.

Klein, G. (1998). *Sources of power: How people make decisions*. MIT Press.

Klein, G. A., Calderwood, R., & Clinton-Cirocco, A. (1986). Rapid decision making on the fire ground. *Proceedings of the Human Factors Society Annual Meeting*, *30*, 576–580. https://doi.org/10.1177/154193128603000616

Komite Nasional Keselamatan Transportasi (KNKT). (2019). *Final report: PT. Lion Mentari Airlines Boeing 737-8 (MAX),*PK-LQP Tanjung Karawang, West Java, Republic of Indonesia 29 October 2018. Republic of Indonesia.

Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind*. University of Chicago Press.

Landman, A., Groen, E. L., van Paassen, M. M. R., Bronkhorst, A. W., & Mulder, M. (2017). Dealing with unexpected events on the flight deck: A conceptual model of startle and surprise. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *59*, 1161–1172. https://doi.org/10.1177/0018720817723428

Landman, A., van Oorschot, P., van Paassen, M. M. R., Groen, E. L., Bronkhorst, A. W., & Mulder, M. (2018). Training pilots for unexpected events: A simulator study on the advantage of unpredictable and variable scenarios. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *60*, 793–805. https://doi.org/10.1177/0018720818779928

Lei, Y., Li, F., Long, C., Li, P., Chen, Q., Ni, Y., & Li, H. (2010). How does typicality of category members affect the deductive reasoning? An ERP study. *Experimental Brain Research*, *204*, 47–56. https://doi.org/10.1007/s00221-010-2292-5

Markman, A. B., & Ross, B. H. (2003). Category use and category learning. *Psychological Bulletin*, *129*, 592–613. https://doi.org/10.1037/0033-2909.129.4.592

Murphy, G. L., & Ross, B. H. (2005). The two faces of typicality in category-based induction. *Cognition*, *95*, 175–200. https://doi.org/10.1016/j.cognition.2004.01.009

National Transportation Safety Board (NTSB). (2014a). *Crash during a nighttime nonprecision instrument approach to landing UPS Flight 1354 Airbus A300-600,* N155UP Birmingham, Alabama August 14, 2013.

National Transportation Safety Board (NTSB). (2014b). *Descent below visual glidepath and impact with Seawall, Asiana Airlines Flight 214, Boeing 777-200ER,* HL7742, San Francisco, California, July 6, 2013.

Nosofsky, R. M. (1988). Similarity, frequency, and category representations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 54–65. https://doi.org/10.1037/0278-7393.14.1.54

Pothos, E. M., & Wills, A. J. (2011). Introduction. In E. M. Pothos & A. J. Wills (Eds.), *Formal approaches in categorization* (pp. 1–17). Cambridge University Press.

Rosch, E. (1975). Cognitive reference points. *Cognitive Psychology*, *7*, 532–547. https://doi.org/10.1016/0010-0285(75)90021-3

Rosch, E. (1978). Principles of categorization. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 27–48). Erlbaum.

Rosch, E., Simpson, C., & Miller, R. S. (1976). Structural bases of typicality effects. *Journal of Experimental Psychology: Human Perception and Performance*, *2*, 491–502. https://doi.org/10.1037/0096-1523.2.4.491

Rossiter, C., & Best, W. (2013). "Penguins don't fly": An investigation into the effect of typicality on picture naming in people with aphasia. *Aphasiology*, *27*, 784–798. https://doi.org/10.1080/02687038.2012.751579

Rothbart, M., Sriram, N., & Davis-Stitt, C. (1996). The retrieval of typical and atypical category members. *Journal of Experimental Social Psychology*, *32*, 309–336. https://doi.org/10.1006/jesp.1996.0015

Sandberg, C., Sebastian, R., & Kiran, S. (2012). Typicality mediates performance during category verification in both ad-hoc and well-defined categories. *Journal of Communication Disorders*, *45*, 69–83. https://doi.org/10.1016/j.jcomdis.2011.12.004

Smith, E. E., & Medin, D. L. (1981). *Categories and concepts* (Vol. 4). Harvard University Press.

Storms, G., De Boeck, P., & Ruts, W. (2000). Prototype and exemplar-based information in natural language categories. *Journal of Memory and Language*, *42*, 51–73. https://doi.org/10.1006/jmla.1999.2669

Wang, X., Tao, Y., Tempel, T., Xu, Y., Li, S., Tian, Y., & Li, H. (2016). Categorization method affects the typicality effect: ERP evidence from a category-inference task. *Frontiers in Psychology*, *7*, 1–11. https://doi.org/10.3389/fpsyg.2016.00184

Weick, K. (1995). *Sensemaking in organizations*. Sage.

Richard Clewley is a doctoral researcher in cognition and flight crew behavior at Cranfield University, UK. He received his MS in ergonomics and organizational behavior from the University of Derby, UK, in 2013.

Jim Nixon is a senior lecturer (associate professor) in human factors at Cranfield University, UK. He received his PhD in human factors from the University of Nottingham, UK, in 2008.

Date received: November 25, 2019
Date accepted: August 29, 2020



https://dspace.org/

Penguins, birds, and pilot knowledge: can an overlooked attribute of human cognition explain our most puzzling aircraft accidents?

Clewley, Richard; Nixon, Jim

2020-10-06
Sage
https://doi.org/10.1177/0018720820960877
Clewley R, Nixon J. (2022) Penguins, birds, and pilot knowledge: can an overlooked attribute of human cognition explain our most puzzling aircraft accidents? Human Factors, Volume 64, Issue 4, June 2022, pp. 662-674

Downloaded from a DSpace repository. https://dspace.org/